D470–D478 *Nucleic Acids Research, 2015, Vol. 43, Database issue*   *Published online 26 November 2014*
doi: 10.1093/nar/gku1204

The BioGRID interaction database: 2015 update

Andrew Chatr-aryamontri[1], Bobby-Joe Breitkreutz[2], Rose Oughtred[3], Lorrie Boucher[2], Sven Heinicke[3], Daici Chen[1], Chris Stark[2], Ashton Breitkreutz[2], Nadine Kolas[2], Lara O'Donnell[2], Teresa Reguly[2], Julie Nixon[4], Lindsay Ramage[4], Andrew Winter[4], Adnane Sellam[5], Christie Chang[3], Jodi Hirschman[3], Chandra Theesfeld[3], Jennifer Rust[3], Michael S. Livstone[3], Kara Dolinski[3] and Mike Tyers[1,2,4,*]

[1]Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Quebec H3C 3J7, Canada, [2]The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada, [3]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA, [4]School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JR, UK and [5]Centre Hospitalier de l'Université Laval (CHUL), Québec, Québec G1V 4G2, Canada

ABSTRACT

The Biological General Repository for Interaction Datasets (BioGRID: http://thebiogrid.org) is an open access database that houses genetic and protein interactions curated from the primary biomedical literature for all major model organism species and humans. As of September 2014, the BioGRID contains 749 912 interactions as drawn from 43 149 publications that represent 30 model organisms. This interaction count represents a 50% increase compared to our previous 2013 BioGRID update. BioGRID data are freely distributed through partner model organism databases and meta-databases and are directly downloadable in a variety of formats. In addition to general curation of the published literature for the major model species, BioGRID undertakes themed curation projects in areas of particular relevance for biomedical sciences, such as the ubiquitin-proteasome system and various human disease-associated interaction networks. BioGRID curation is coordinated through an Interaction Management System (IMS) that facilitates the compilation interaction records through structured evidence codes, phenotype ontologies, and gene annotation. The BioGRID architecture has been improved in order to support a broader range of interaction and post-translational modification types, to allow the representation of more complex multi-gene/protein interactions, to account for cellular phenotypes through structured ontologies, to expedite curation through semi-automated text-mining approaches, and to enhance curation quality control.

INTRODUCTION

Massive increases in high-throughput DNA sequencing technologies (1) have enabled an unprecedented level of genome annotation for many hundreds of species (2–6), which has led to tremendous progress in the understanding of gene organization, genome evolution and the genetic basis for disease. At the same time, sequencing-based methods have uncovered many intricacies of gene regulation at a genomic scale, including expression patterns, alternative splicing, non-coding transcription and the myriad of regulatory factors that bind DNA and RNA (7–11). Proteomics approaches, largely based on mass spectrometry, have similarly mapped the abundance and post-translational modifications of proteins at impressive depth of coverage (12–15). At the phenotypic level, genome-wide reagent collections for systematic perturbation of gene function have led to compendia of functional profiles for many different phenotypic characteristics (16–19). This wealth of new data has been accrued in model organism systems, and particularly in humans, in both normal and disease contexts. In spite of this data deluge, the fundamental problem of how genotype is translated into phenotype, and how genetic mutations can affect this complex relationship, remains a formidable roadblock in our understanding of fundamental biology and the basis for human disease.

It is now evident that genes and their encoded proteins function in the context of a vast, dynamic network of interactions (20–23). The generation of comprehensive genetic and protein interaction maps will thus be essential for unraveling the many complexities of biological processes and for understanding the general genotype to phenotype map-

---

ping problem (24). For example, the integration of genetic interaction networks with other genome-wide data types has helped to explain how sets of genes function differently in specific cellular contexts, conditions or tissues (25–28). The systematic experimental identification and characterization of protein and genetic interaction networks in major model organism species and humans has continued to grow in pace and scale (21,23,29–35). With such interaction datasets in hand, it has been possible to implement computational methods for analysis and prediction of the response of cellular networks to perturbation by disease-associated mutation or pathogen infection (28,36–38).

The comprehensive annotation and compilation of all known biological interactions in a computable form is essential for network-based approaches to understanding biological systems and human disease (39). The Biological General Repository for Interaction Datasets (BioGRID: http://thebiogrid.org) was established in order to help capture biological interaction data from the primary biomedical literature and to provide this data in a readily computable format (40). BioGRID collects and annotates genetic and protein interaction data from the published literature for all major model organism species and humans. When available, data on the influence of protein post-translational modifications, including phosphorylation and ubiquitination, is also captured. The complete BioGRID dataset is freely accessible through a dedicated web-based search portal and is also available for download in various standardized formats. BioGRID data content is updated and permanently archived on a monthly basis, and in addition to the BioGRID web interface, is disseminated to the research community through model organism database (MOD) partners (41–46) and other biological resources and meta-databases (47–52). The interaction datasets in BioGRID thus provide a resource for biomedical researchers who study the function of individual genes and pathways, as well as for computational biologists who analyze the properties of large biological networks.

DATABASE GROWTH AND STATISTICS

The current BioGRID release (August 2014, version 3.2.115) houses a total of 749 912 interactions (515 032 non-redundant) comprising 471 525 protein (physical) interactions (318 069 non-redundant) and 278 387 genetic interactions (204 801 non-redundant) (Table 1). The number of interactions housed in BioGRID has increased by ∼50% since the 2013 BioGRID update (40). All data in BioGRID has been manually curated from a total of 43 149 articles indexed in PubMed (Figure 1). BioGRID also currently contains data on 42 907 protein phosphorylation sites, which are mainly drawn from high-throughput mass spectrometry studies, as housed in the PhosphoGRID database (53). In 2014, Google Analytics reported that the BioGRID received on average 88 080 page views and 12 399 unique visitors per month, versus 69 237 page views and 10 110 unique visitors per month in 2012. BioGRID data files were downloaded on average 9256 times per month in 2014, compared with 6900 downloads per month in 2012. These statistics do not include the widespread dissemination of BioGRID records by various partner databases and meta-resources.

In 2014, the BioGRID user base was located primarily in the USA (30%), followed by China (8%), United Kingdom (7%), Canada (6%), Germany (6%), Japan (6%), India (4%), France (4%), Spain (2%) and all other countries (27%).

DATA CURATION AND QUALITY CONTROL

BioGRID continues to maintain complete curation of the primary literature for genetic and protein interactions in the model yeasts *Saccharomyces cerevisiae* (342 878 total interactions) and *Schizosaccharomices pombe* (68 015 total interactions). These datasets are updated on a monthly basis and released for redistribution through the *Saccharomyces* Genome Database (41) and PomBase (43). In addition to these two yeasts, BioGRID contains interaction data for more than 30 model organisms at varying depths of coverage. However, the immense extent of the biomedical literature—more than 24 million articles in PubMed as of August 2014—and its ever-accelerating rate of growth render the complete manual curation of all interaction data virtually impossible (39). The identification of publications that contain actual interaction data is a non-trivial step in the curation workflow (54). Although the entire BioGRID dataset is drawn directly from just 43 149 publications, in reality several-fold more publications have been directly parsed by curators, usually in an entirely manual fashion (55). While our initial strategy for the identification of relevant papers was based on simple PubMed searches based on keywords and/or gene names, we now prioritize literature queues for different projects through advanced text-mining approaches. For example, BioGRID has several projects that are facilitated by Support Vector Machine (SVM) analyses carried out in collaboration with the Textpresso text-mining group (56). We have also begun to use text-mining for the curation of protein phosphorylation sites through a collaboration with developers of the RLIMS-P system (57). To facilitate the development of improved text-mining approaches, the BioGRID routinely contributes to the BioCreative (Critical Assessment of Information Extraction in Biology) challenge by providing test datasets and curation expertise (58–60).

Curation accuracy and consistency are critical for the integrity of the BioGRID resource. The Interaction Management System (IMS) that is used to coordinate curation efforts helps ensure that only unambiguous and appropriate gene identifiers are used. For direct submission of high-throughput datasets to BioGRID, curators work closely with data providers to ensure proper data representation, particularly for quantitative datasets. For example, BioGRID recently incorporated a pre-publication dataset of 23 756 human protein interactions detected by quantitative affinity capture-mass spectrometry (35), as generated by the Gygi and Harper groups. BioGRID also provides an e-mail based helpdesk for evaluation and correction of dubious entries noticed by authors or other users. Importantly, as each monthly BioGRID update is permanently archived, users are able to trace any alterations to the dataset, and thereby easily assess any potential impact on analyses that may have been performed. BioGRID has also recently implemented an automated random re-curation procedure, whereby small subsets of interactions derived from low-

Table 1. Increase in BioGRID data content

Organism	Type	August 2012 (3.1.92)			August 2014 (3.2.115)		
		Nodes	Edges	Publications	Nodes	Edges	Publications
Arabidopsis thaliana	PI	5915	16 476	1118	7200	21 536	1414
	GI	107	188	62	112	192	66
Caenorhabditis elegans	PI	2927	5010	93	3288	6345	178
	GI	1109	2326	22	1129	2344	30
Drosophila melanogaster	PI	7998	35 843	314	8076	37 606	416
	GI	1023	9934	1468	1042	9980	1483
Homo sapiens	PI	14 896	123 436	17 134	18 435	237 498	23 388
	GI	1291	1609	237	1364	1678	273
S. cerevisiae	PI	6003	114 506	6601	6410	135 690	7402
	GI	5561	189 692	6686	5674	207 188	7257
S. Pombe	PI	1773	6019	968	2694	11 270	1146
	GI	1907	14 015	1158	3158	56 745	1359
Other organisms	ALL	8435	15 978	2724	16470	35 347	5269
Total	ALL	44 515	527 569	33 858	55 528	749 912	43 149

Data drawn from monthly release 3.1.92 and 3.2.115 of BioGRID.
Nodes refers to genes or proteins, edges refers to interactions.
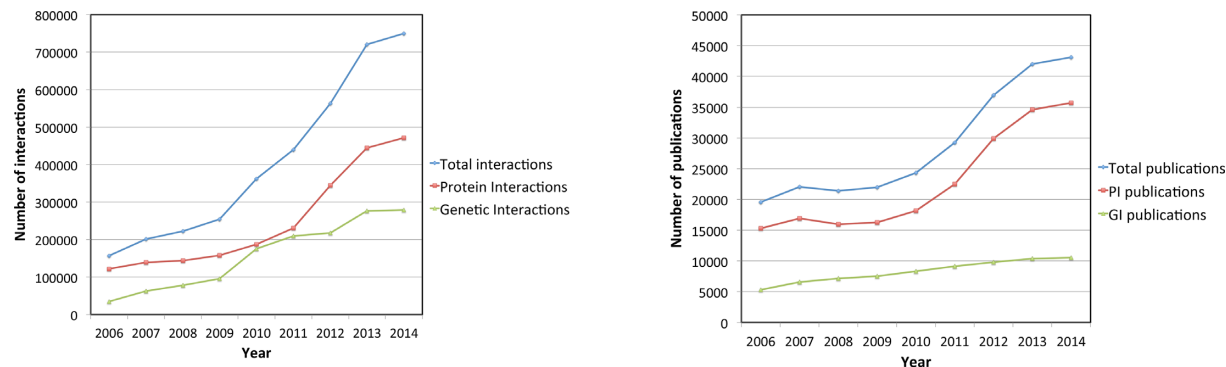PI, protein (physical) interactions; GI, genetic interactions.

Figure 1. Growth of the BioGRID database. Increments in interaction records and source publications reported in BioGRID from July 2006 (release 2.0.18) to August 2014 (release 3.2.115). Left panel shows the increase of annotated protein interactions (PI, red), genetic interactions (GI, green) and total interactions (blue). Right panel shows the number of publications that report protein or genetic interactions and the total number of curated publications.

throughput studies are blindly re-curated in order to ensure curation consistency.

THEMED CURATION PROJECTS

To maximize depth of BioGRID curation coverage in specific areas relevant to human disease, we have undertaken a series of themed curation projects delineated by a specific biological process or a specific disease topic. These themed curation efforts are implemented in three discrete steps: (i) compilation of a structured gene annotation reference list for the project, typically in consultation with domain experts; (ii) generation of a list of all candidate publications through custom PubMed queries and text-mining approaches and; (iii) curation of the interaction data according to structured evidence codes as coordinated through the automated IMS curation interface. In the largest such project to date, we have curated the entire literature for interactions associated with the ubiquitin-proteasome system (UPS). Manual expert compilation of a comprehensive UPS gene reference list was augmented by semi-automated parsing of protein domain and protein function annotations available through a number of sequence-based databases

(48,49,61–63). We thus annotated 1251 human genes to the UPS in a structured format that classified each gene according to enzymatic and other functional characteristics. This gene list was then used to seed PubMed searches to generate a prioritized curation queue of ∼20 000 publications. As will be reported elsewhere in detail, a sustained manual curation effort allowed the construction of a dataset of 102 906 interactions (50 561 non redundant) in the human UPS. In addition, we carried out the systematic annotation of ubiquitination sites detected by high-throughput mass-spectrometry-based approaches (64–66).

A second major curation theme undertaken recently at BioGRID is the arachidonic acid pathway (AAP) as part of the Personalized NSAID Therapeutics Consortium (PEN-TACON) project (http://www.pentaconhq.org). The AAP is the primary cellular mechanism for production of pain and inflammation mediators, and is also involved in renal function and homeostasis (67) Core genes involved in the AAP, as well as AAP-related genes and genes involved in blood pressure (BP) regulation, were identified using curated pathway resources such as KEGG (68) and Reactome (69), as well as Gene Ontology (63) annotations. These gene lists

were further expanded via on-going literature review and by input from domain experts associated with the PEN-TACON project. BioGRID curators directly reviewed over 2400 papers and curated more than 1300 AAP protein interactions, 49% of which were from low-throughput studies. This curation effort was then broadened to include AAP-related and BP-related proteins to yield an additional 1200 interactions (84% low-throughput) and 2100 interactions (70% low-throughput), respectively.

Each themed project will be associated with a specific project page in the BioGRID web interface, which will enable users to identify and query specific gene lists within each project. Similarly, project-specific download datasets will be made available and updated on a monthly basis. Other themed curation areas in progress include projects on Parkinson's Disease (PD) and other neurobiological disorders, breast cancer, the Wnt signaling pathway, the chromatin modification system, the autophagy system and ubiquitin-like modifiers. We encourage enquiries from potential expert collaborators with an interest in interaction curation projects with a particular focus on a human disease or a conserved biological process.

DATA STANDARDS

BioGRID curation is based on a structured but simplified set of experimental evidence codes for the representation of protein (physical) and genetic interactions. The BioGRID data model allows for the representation of both binary and higher order interactions. BioGRID evidence codes map directly to the Molecular Interaction Ontology, which is maintained by the Proteomics Standards Initiative (70), thereby making BioGRID data records fully interoperable with other datasets released in PSI-MI format. BioGRID evidence codes are periodically updated to reflect new advances in experimental methods. For instance, a Proximity Label-Mass Spectrometry (MS) evidence code was recently introduced in order to document interactions detected upon covalent modification of interaction partners by diffusible reactive species produced by a bait-enzyme fusion protein (71). All evidence codes are fully documented on the BioGRID help wiki section (http://wiki.thebiogrid.org/doku.php/experimental_systems).

BioGRID has recently collaborated with WormBase (45) to develop a new Genetic Interaction (GI) Ontology. This standard has been approved by the main MODs, including SGD (41), CGD (72), PomBase (43), ZFIN (46), FlyBase (42) and TAIR (73). The new GI ontology reconciles different terminologies often used by the biomedical research community and across different MODs. The GI Ontology is based on a previous standard (74), but extends the list of GI terms and inequalities to provide more granular terms based on terminology that is familiar to geneticists (75,76). These GI terms are structured in an ontological format whereby the relationships between the various interaction types are precisely defined. The GI ontology is also available in a simplified slim version of only 23 terms that cover the majority of the genetic interaction cases curated by various MODs. These newly standardized GI terms will facilitate the interpretation of genetic interactions, enable the integration of large genetic interaction datasets, and allow cross-species comparisons of genetic interaction networks. We note that BioGRID currently contains 265 000 yeast genetic interactions associated with over 600 unique phenotypes, which will be automatically remapped to the new GI ontology terms in future releases. The GI ontology is now available as part of the Proteomics Standards Initiative-Molecular Interaction (PSI-MI) ontology (70) and will be published in full in the near future (Grove *et al.*, in preparation).

DATABASE IMPROVEMENTS

The web-based IMS curation interface for the BioGRID has recently undergone major revisions in order to allow more sophisticated annotation for future curation projects. The IMS core architecture now enables curation of a broader range of interaction types including for proteins, genes, RNA, small molecules, domains and protein fragments. The overall database architecture has also been improved to allow representation of higher order relationships between interacting partners, such as triple mutant combinations, protein complexes, chemical-genetic interactions and post-translational modifications (Figure 2). The IMS has been elaborated to include more than a dozen comprehensive new ontologies (77–79) that allow curators to unambiguously record new details of any relationship, such as cell lines, phenotypes, small molecules, alleles, diseases, tissues and enzymes (Figure 3). IMS features for curation tracking, fault tolerance and overall curation quality have also been improved. For example, to accommodate more frequent deposition of high-throughput datasets in BioGRID, new tracking tools enable the long-term storage of Supplementary Data files for archival and data reconciliation purposes. The IMS can also track the decision-making processes of each curator for each specific publication, such that it is possible to trace decisions even when the original source material is no longer available or the curator is no longer a member of the BioGRID team. To improve the overall fault tolerance of the underlying database architecture, we have continuously updated our MySQL database platform to utilize enhancements such as InnoDB tables and transactional logging.

The BioGRID is currently deployed on five virtual machines (VMs) hosted by a commercial third party provider. The VMs are fully customizable and provide state-of-the-art Intel Ivy Bridge processors, application-specific memory that is scalable from 1 to 96 GB and industry-leading native SSD high performance storage that can be readily expanded as needed. Each system has a fully redundant backup that runs daily and weekly and is situated on a 40 GB network that allows for fast access by BioGRID developers and curators in different countries, as well as by web interface and REST service users. Since deployment to cloud-based servers two years ago, the BioGRID has maintained at least 99.9% uptime, without a single major system failure. Each deployment is routinely refreshed with new hardware and software updates that keep pace with changing requirements, demands for higher usage, and system stability and security.

The IMS and the BioGRID have been improved through a new comprehensive annotation system. Our previous sys-
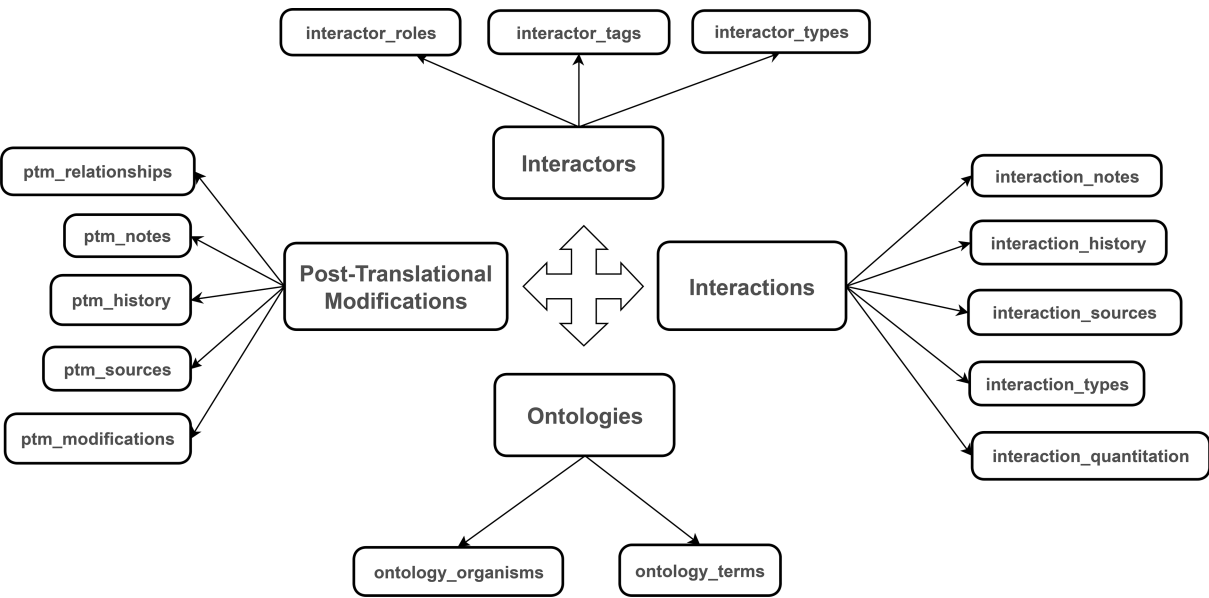
Figure 2. The Interaction Management System. Overview of the new database architecture that allows BioGRID to transition from a pairwise interaction format to an n-way interaction format for representation of complex protein or genetic interaction relationships. The database schema has also been extended to include support for post-translational modification (PTM) and phenotype curation. The central components illustrated (Interactors, Post-translational Modifications, Interactions, and Ontologies) represent the four major sectors of the IMS architecture. Partial representations of the child support tables that link to the main parent tables are shown but precise entity relationships are not indicated. In total, the IMS contains 57 interlinked tables. All controlled vocabularies (experimental systems, modifications and tags) have been converted into formal ontologies in order to remove redundancies present in the previous database architecture.
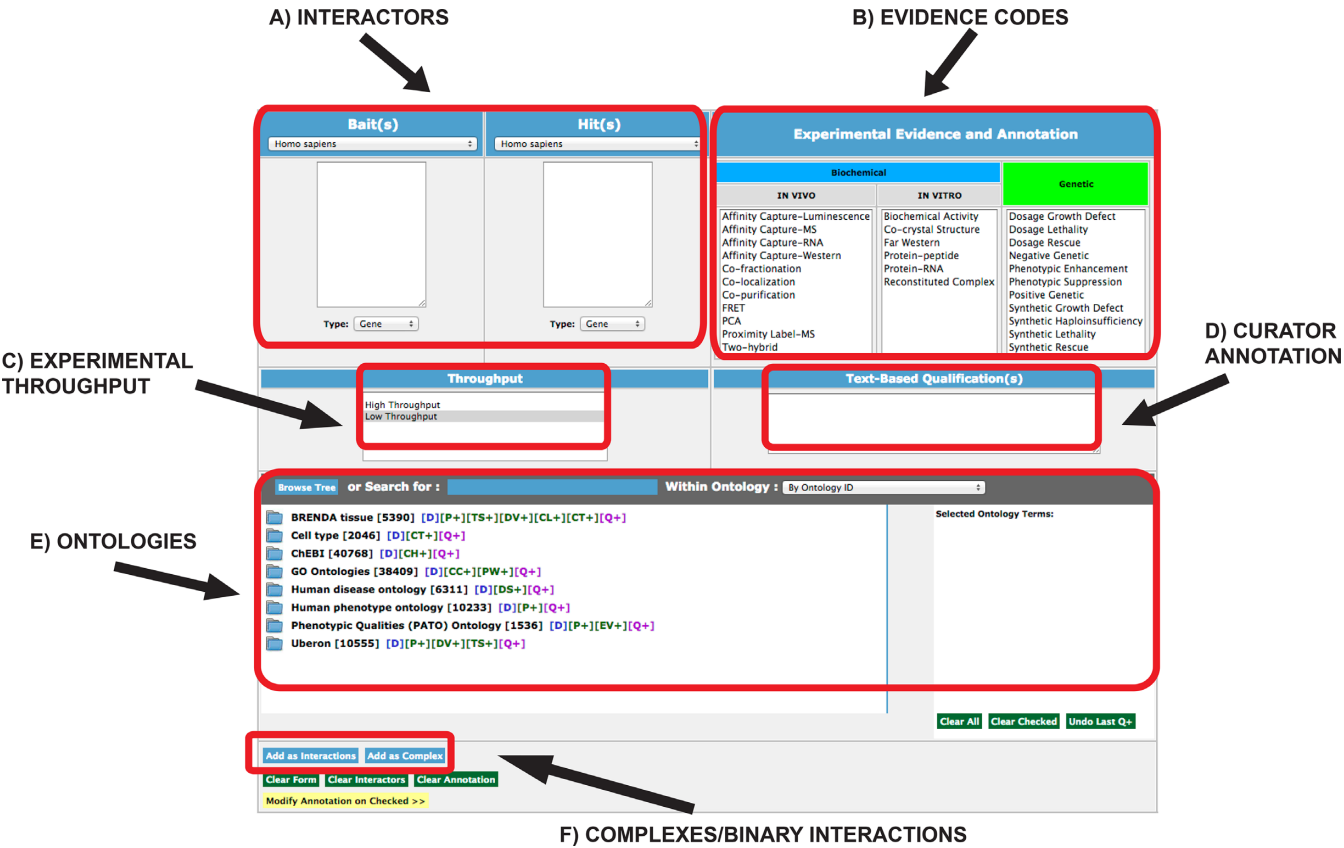
Figure 3. Snapshot of the new IMS curation interface. The main functionalities available to BioGRID curators in the new IMS for the annotation of protein and genetic interactions are shown (**A**–**D**). The new system is based on ontologies (**E**) for the annotation of gene function (Gene Ontology), cell type (Cell Type Ontology), tissue (BRENDA Tissue ontology), small molecules (CheBI), human disease (Human Disease Ontology), human phenotypes (Human Phenotype Ontology, Phenotypic Qualities Ontologies) or anatomical structures (Uberon) and accepts annotation of binary and n-way interactions (**F**).

tem included more than 28 million unique aliases, identifiers, systematic names and MOD references for over 100 supported model organisms. The updated annotation platform provides 20 million additional references and support for many additional organisms. The new data records will allow faster curation as obscure identifiers used in older publications can be easily translated into common references that are recognizable by most major MODs. The local storage of annotations in the new system also improves robustness of the internal curation pipeline by obviating the need for external APIs. The annotation system is updated on a regular basis and allows for straightforward incorporation of new organisms and facile adaptation to major annotation changes. These enhancements to the database architecture maximize performance and flexibility in curation tasks, especially for HTP datasets.

DATA DISSEMINATION

All BioGRID datasets and interaction records can be accessed and interrogated by a variety of different means. The BioGRID web page allows searches of interaction data by gene name, gene aliases or PMID publication identifiers. The complete BioGRID dataset or subsets thereof are also available for download in a number of tabular (tab, tab2 and mitab) and XML (PSI-MI 1.0, PSI-MI 2.5) formats. A detailed step-by-step guide to the BioGRID web interface is now available (Oughtred *et al.*, submitted). BioGRID interaction data is also accessible to the individual researcher indirectly through a number of other biological databases including NCBI Entrez-Gene (48), Uniprot (49), DroID (80), GermOnline (81), FlyBase (42), TAIR (73), SGD (41), PomBase (43), STRING (47), iRefIndex (82), GeneMania (83) and Pathway Commons (50).

Software developers can access BioGRID data directly through the BioGRID representational state transfer (REST) service (84). The BioGRID Webgraph (84) and the BioGRID Cytoscape plugin (84) utilize the REST service for the visualization and analysis of BioGRID interaction networks. The BioGRID REST service application program interface (API) has been completely rebuilt to improve performance, enhance reliability and support scalability through more powerful server hardware available in the cloud. This transition to cloud-based servers has reduced query response times from an average of 5.1 s to <0.02 ms. As a direct result of these changes, the BioGRID REST service now supports more than 350 worldwide active projects that perform more than 100 000 queries per month with an average return of more than 2 million interactions per month. For example, the ProHits open source mass spectrometry LIMS platform uses the REST service to incorporate BioGRID data into analysis of experimental mass spectrometry data (85,86). The BioGRID Cytoscape plugin version 2.3 has also been redesigned to take advantage of the improvements made to the REST API and can be downloaded directly from the BioGRID website at http://wiki.thebiogrid.org/doku.php/tools. Finally, we have also implemented support for the PSICQUIC API interface (87), which has resulted in more than 140 000 queries per month from a wide variety of users.

FUTURE DEVELOPMENTS

The BioGRID will continue its core mandate to curate biological interaction data from the primary biomedical literature across the major model organism species and humans for unrestricted dissemination to the research community. The BioGRID database architecture will continue to be improved through additional updates to the IMS curation management system that will facilitate the routine deposition of pre-publication large-scale quantitative datasets, allow the capture of detailed phenotype information associated with genetic interactions, and further extend the internal annotation system to new organisms. Future themed curation drives will be focused on conserved biological processes such as the autophagy system and specific human diseases such as neurological and cardiac disorders. All interaction data for themed projects will be made accessible through project-specific web interfaces. The BioGRID will also continue to exploit text-mining technologies in order to improve the efficiency of curation workflows for future themed projects. BioGRID curation parameters for these projects will be extended to additional post-translational modifications, context-specific effects and structured phenotypes. New computational approaches based on integration of genome-scale datasets will be used to develop tissue- and disease-specific functional networks that will help guide and validate expert manual curation. This disease network-associated curation will be augmented through the capture of relevant drug or small molecule interactions. Collectively, these approaches will enable efficient cross-species comparisons of biological interaction networks, particularly for identification of new models of human disease.

REFERENCES

1. Shendure,J. and Lieberman Aiden,E. (2012) The expanding scsope of DNA sequencing. *Nat. Biotechnol.*, **30**, 1084–1094.
2. Genomes Project Consortium, Abecasis,G.R., Auton,A., Brooks,L.D., DePristo,M.A., Durbin,R.M., Handsaker,R.E., Kang,H.M., Marth,G.T. and McVean,G.A. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, **491**, 56–65.
3. Miquel,S., Peyretaillade,E., Claret,L., de Vallee,A., Dossat,C., Vacherie,B., Zineb el,H., Segurens,B., Barbe,V., Sauvanet,P. *et al.* (2010) Complete genome sequence of Crohn's disease-associated adherent-invasive E. coli strain LF82. *PLoS One*, **5**, e12714.
4. Mouse Genome Sequencing Consortium, Waterston,R.H., Lindblad-Toh,K., Birney,E., Rogers,J., Abril,J.F., Agarwal,P., Agarwala,R., Ainscough,R., Alexandersson,M. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.
5. Pontius,J.U., Mullikin,J.C., Smith,D.R., Agencourt Sequencing,T., Lindblad-Toh,K., Gnerre,S., Clamp,M., Chang,J., Stephens,R., Neelam,B. *et al.* (2007) Initial sequence and comparative analysis of the cat genome. *Genome Res.*, **17**, 1675–1689.
6. Prufer,K., Racimo,F., Patterson,N., Jay,F., Sankararaman,S., Sawyer,S., Heinze,A., Renaud,G., Sudmant,P.H., de Filippo,C. *et al.* (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, **505**, 43–49.
7. Parada,G.E., Munita,R., Cerda,C.A. and Gysling,K. (2014) A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic Acids Res.*, **42**, 10564–10578.
8. Lappalainen,T., Sammeth,M., Friedlander,M.R., 't Hoen,P.A., Monlong,J., Rivas,M.A., Gonzalez-Porta,M., Kurbatova,N., Griebel,T., Ferreira,P.G. *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506–511.
9. Kellis,M., Wold,B., Snyder,M.P., Bernstein,B.E., Kundaje,A., Marinov,G.K., Ward,L.D., Birney,E., Crawford,G.E., Dekker,J. *et al.* (2014) Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 6131–6138.
10. Braunschweig,U., Gueroussov,S., Plocik,A.M., Graveley,B.R. and Blencowe,B.J. (2013) Dynamic integration of splicing within gene regulatory pathways. *Cell*, **152**, 1252–1269.
11. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
12. Choudhary,C., Kumar,C., Gnad,F., Nielsen,M.L., Rehman,M., Walther,T.C., Olsen,J.V. and Mann,M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, **325**, 834–840.
13. Hendriks,I.A., D'Souza,R.C., Yang,B., Verlaan-de Vries,M., Mann,M. and Vertegaal,A.C. (2014) Uncovering global SUMOylation signaling networks in a site-specific manner. Nat. Struct. Mol. Biol., **21**, 927–936.
14. Sharma,K., D'Souza,R.C., Tyanova,S., Schaab,C., Wisniewski,J.R., Cox,J. and Mann,M. (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.*, **8**, 1583–1594.
15. Udeshi,N.D., Mani,D.R., Eisenhaure,T., Mertins,P., Jaffe,J.D., Clauser,K.R., Hacohen,N. and Carr,S.A. (2012) Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition. *Mol. Cell. Proteomics*, **11**, 148–159.
16. Neumann,B., Walter,T., Heriche,J.K., Bulkescher,J., Erfle,H., Conrad,C., Rogers,P., Poser,I., Held,M., Liebel,U. *et al.* (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature*, **464**, 721–727.
17. Deshpande,R., Asiedu,M.K., Klebig,M., Sutor,S., Kuzmin,E., Nelson,J., Piotrowski,J., Shin,S.H., Yoshida,M., Costanzo,M. *et al.* (2013) A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res.*, **73**, 6128–6136.
18. Nichols,R.J., Sen,S., Choo,Y.J., Beltrao,P., Zietek,M., Chaba,R., Lee,S., Kazmierczak,K.M., Lee,K.J., Wong,A. *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell*, **144**, 143–156.
19. Carette,J.E., Guimaraes,C.P., Wuethrich,I., Blomen,V.A., Varadarajan,M., Sun,C., Bell,G., Yuan,B., Muellner,M.K.,
20. Nijman,S.M. *et al.* (2011) Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.*, **29**, 542–546.
20. Rigbolt,K.T., Prokhorova,T.A., Akimov,V., Henningsen,J., Johansen,P.T., Kratchmarova,I., Kassem,M., Mann,M., Olsen,J.V. and Blagoev,B. (2011) System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal*, **4**, rs3.
21. Breitkreutz,A., Choi,H., Sharom,J.R., Boucher,L., Neduva,V., Larsen,B., Lin,Z.Y., Breitkreutz,B.J., Stark,C., Liu,G. *et al.* (2010) A global protein kinase and phosphatase interaction network in yeast. *Science*, **328**, 1043–1046.
22. Vidal,M., Cusick,M.E. and Barabasi,A.L. (2011) Interactome networks and human disease. *Cell*, **144**, 986–998.
23. Costanzo,M., Baryshnikova,A., Bellay,J., Kim,Y., Spear,E.D., Sevier,C.S., Ding,H., Koh,J.L., Toufighi,K., Mostafavi,S. *et al.* (2010) The genetic landscape of a cell. *Science*, **327**, 425–431.
24. Beltrao,P., Ryan,C. and Krogan,N.J. (2012) Comparative interaction networks: bridging genotype to phenotype. *Adv. Exp. Med. Biol.*, **751**, 139–156.
25. Lundby,A., Rossin,E.J., Steffensen,A.B., Acha,M.R., Newton-Cheh,C., Pfeufer,A., Lynch,S.N., Consortium,Q.T. I.I. G., Olesen,S.P., Brunak,S. *et al.* (2014) Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. *Nat. Methods*, **11**, 868–874.
26. Lage,K., Greenway,S.C., Rosenfeld,J.A., Wakimoto,H., Gorham,J.M., Segre,A.V., Roberts,A.E., Smoot,L.B., Pu,W.T., Pereira,A.C. *et al.* (2012) Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 14035–14040.
27. Lage,K., Mollgard,K., Greenway,S., Wakimoto,H., Gorham,J.M., Workman,C.T., Bendsen,E., Hansen,N.T., Rigina,O., Roque,F.S. *et al.* (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Mol. Syst. Biol.*, **6**, 381.
28. Guan,Y., Gorenshteyn,D., Burmeister,M., Wong,A.K., Schimenti,J.C., Handel,M.A., Bult,C.J., Hibbs,M.A. and Troyanskaya,O.G. (2012) Tissue-specific functional networks for prioritizing phenotype and disease genes. *PLoS Comput. Biol.*, **8**, e1002694.
29. Rajagopala,S.V., Sikorski,P., Kumar,A., Mosca,R., Vlasblom,J., Arnold,R., Franca-Koh,J., Pakala,S.B., Phanse,S., Ceol,A. *et al.* (2014) The binary protein-protein interaction landscape of Escherichia coli. *Nat. Biotechnol.*, **32**, 285–290.
30. Babu,M., Vlasblom,J., Pu,S., Guo,X., Graham,C., Bean,B.D., Burston,H.E., Vizeacoumar,F.J., Snider,J., Phanse,S. *et al.* (2012) Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. *Nature*, **489**, 585–589.
31. Havugimana,P.C., Hart,G.T., Nepusz,T., Yang,H., Turinsky,A.L., Li,Z., Wang,P.I., Boutz,D.R., Fong,V., Phanse,S. *et al.* (2012) A census of human soluble protein complexes. *Cell*, **150**, 1068–1081.
32. Babu,M., Diaz-Mejia,J.J., Vlasblom,J., Gagarinova,A., Phanse,S., Graham,C., Yousif,F., Ding,H., Xiong,X., Nazarians-Armavil,A. *et al.* (2011) Genetic interaction maps in Escherichia coli reveal functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet.*, **7**, e1002377.
33. Corominas,R., Yang,X., Lin,G.N., Kang,S., Shen,Y., Ghamsari,L., Broly,M., Rodriguez,M., Tam,S., Trigg,S.A. *et al.* (2014) Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat. Commun.*, **5**, 3650.
34. Rozenblatt-Rosen,O., Deo,R.C., Padi,M., Adelmant,G., Calderwood,M.A., Rolland,T., Grace,M., Dricot,A., Askenazi,M., Tavares,M. *et al.* (2012) Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. *Nature*, **487**, 491–495.
35. Sowa,M.E., Bennett,E.J., Gygi,S.P. and Harper,J.W. (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell*, **138**, 389–403.
36. Hofree,M., Shen,J.P., Carter,H., Gross,A. and Ideker,T. (2013) Network-based stratification of tumor mutations. *Nat. Methods*, **10**, 1108–1115.
37. Gulbahce,N., Yan,H., Dricot,A., Padi,M., Byrdsong,D., Franchi,R., Lee,D.S., Rozenblatt-Rosen,O., Mar,J.C., Calderwood,M.A. *et al.*

(2012) Viral perturbations of host networks reflect disease etiology. *PLoS Comput. Biol.*, **8**, e1002531.

38. Taylor,I.W., Linding,R., Warde-Farley,D., Liu,Y., Pesquita,C., Faria,D., Bull,S., Pawson,T., Morris,Q. and Wrana,J.L. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.*, **27**, 199–204.

39. Dolinski,K., Chatr-Aryamontri,A. and Tyers,M. (2013) Systematic curation of protein and genetic interaction data for computable biology. *BMC Biol.*, **11**, 43.

40. Chatr-Aryamontri,A., Breitkreutz,B.J., Heinicke,S., Boucher,L., Winter,A., Stark,C., Nixon,J., Ramage,L., Kolas,N., O'Donnell,L. *et al.* (2013) The BioGRID interaction database: 2013 update. *Nucleic Acids Res.*, **41**, D816–D823.

41. Cherry,J.M., Hong,E.L., Amundsen,C., Balakrishnan,R., Binkley,G., Chan,E.T., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R. *et al.* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.

42. St Pierre,S.E., Ponting,L., Stefancsik,R., McQuilton,P. and FlyBase,C. (2014) FlyBase 102–advanced approaches to interrogating FlyBase. *Nucleic Acids Res.*, **42**, D780–D788.

43. Wood,V., Harris,M.A., McDowall,M.D., Rutherford,K., Vaughan,B.W., Staines,D.M., Aslett,M., Lock,A., Bahler,J., Kersey,P.J. *et al.* (2012) PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.*, **40**, D695–D699.

44. Blake,J.A., Bult,C.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Mouse Genome Database,G. (2014) The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res.*, **42**, D810–D817.

45. Harris,T.W., Baran,J., Bieri,T., Cabunoc,A., Chan,J., Chen,W.J., Davis,P., Done,J., Grove,C., Howe,K. *et al.* (2014) WormBase 2014: new views of curated biology. *Nucleic Acids Res.*, **42**, D789–D793.

46. Bradford,Y., Conlin,T., Dunn,N., Fashena,D., Frazer,K., Howe,D.G., Knight,J., Mani,P., Martin,R., Moxon,S.A. *et al.* (2011) ZFIN: enhancements and updates to the Zebrafish Model Organism Database. *Nucleic Acids Res.*, **39**, D822–D829.

47. Franceschini,A., Szklarczyk,D., Frankild,S., Kuhn,M., Simonovic,M., Roth,A., Lin,J., Minguez,P., Bork,P., von Mering,C. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.

48. Ncbi Resource Coordinators (2014) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **42**, D7–D17.

49. UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.*, **42**, D191–D198.

50. Cerami,E.G., Gross,B.E., Demir,E., Rodchenkov,I., Babur,O., Anwar,N., Schultz,N., Bader,G.D. and Sander,C. (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.

51. Orchard,S., Kerrien,S., Abbani,S., Aranda,B., Bhate,J., Bidwell,S., Bridge,A., Briganti,L., Brinkman,F.S., Cesareni,G. *et al.* (2012) Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat. Methods*, **9**, 345–350.

52. Warde-Farley,D., Donaldson,S.L., Comes,O., Zuberi,K., Badrawi,R., Chao,P., Franz,M., Grouios,C., Kazi,F., Lopes,C.T. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.

53. Sadowski,I., Breitkreutz,B.J., Stark,C., Su,T.C., Dahabieh,M., Raithatha,S., Bernhard,W., Oughtred,R., Dolinski,K., Barreto,K. *et al.* (2013) The PhosphoGRID Saccharomyces cerevisiae protein phosphorylation site database: version 2.0 update. *Database*, **2013**, bat026.

54. Hirschman,L., Burns,G.A., Krallinger,M., Arighi,C., Cohen,K.B., Valencia,A., Wu,C.H., Chatr-Aryamontri,A., Dowell,K.G., Huala,E. *et al.* (2012) Text mining for the biocuration workflow. *Database*, **2012**, bas020.

55. Reguly,T., Breitkreutz,A., Boucher,L., Breitkreutz,B.J., Hon,G.C., Myers,C.L., Parsons,A., Friesen,H., Oughtred,R., Tong,A. *et al.* (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J. Biol.*, **5**, 11.

56. Muller,H.M., Kenny,E.E. and Sternberg,P.W. (2004) Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS Biol.*, **2**, e309.

57. Torii,M., Li,G., Li,Z., Oughtred,R., Diella,F., Celen,I., Arighi,C.N., Huang,H., Vijay-Shanker,K. and Wu,C.H. (2014) RLIMS-P: an online text-mining tool for literature-based extraction of protein phosphorylation information. *Database*, **2014**, bau081.

58. Krallinger,M., Vazquez,M., Leitner,F., Salgado,D., Chatr-Aryamontri,A., Winter,A., Perfetto,L., Briganti,L., Licata,L., Iannuccelli,M. *et al.* (2011) The Protein-Protein interaction tasks of BioCreative III: classification/ranking of articles and linking bio-ontology concepts to full text. *BMC Bioinformatics*, **12(Suppl. 8)**, S3.

59. Chatr-Aryamontri,A., Winter,A., Perfetto,L., Briganti,L., Licata,L., Iannuccelli,M., Castagnoli,L., Cesareni,G. and Tyers,M. (2011) Benchmarking of the 2010 BioCreative Challenge III text-mining competition by the BioGRID and MINT interaction databases. *BMC Bioinformatics*, **12(Suppl. 8)**, S8.

60. Kwon,D., Kim,S., Shin,S.Y., Chatr-aryamontri,A. and Wilbur,W.J. (2014) Assisting manual literature curation for protein-protein interactions using BioQRator. *Database*, **2014**, bau067.

61. Hunter,S., Jones,P., Mitchell,A., Apweiler,R., Attwood,T.K., Bateman,A., Bernard,T., Binns,D., Bork,P., Burge,S. *et al.* (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.*, **40**, D306–D312.

62. Finn,R.D., Bateman,A., Clements,J., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Heger,A., Hetherington,K., Holm,L., Mistry,J. *et al.* (2014) Pfam: the protein families database. *Nucleic Acids Res.*, **42**, D222–D230.

63. Gene Ontology,C., Blake,J.A., Dolan,M., Drabkin,H., Hill,D.P., Li,N., Sitnikov,D., Bridges,S., Burgess,S., Buza,T. *et al.* (2013) Gene Ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.

64. Sarraf,S.A., Raman,M., Guarani-Pereira,V., Sowa,M.E., Huttlin,E.L., Gygi,S.P. and Harper,J.W. (2013) Landscape of the PARKIN-dependent ubiquitylome in response to mitochondrial depolarization. *Nature*, **496**, 372–376.

65. Oshikawa,K., Matsumoto,M., Oyamada,K. and Nakayama,K.I. (2012) Proteome-wide identification of ubiquitylation sites by conjugation of engineered lysine-less ubiquitin. *J. Proteome Res.*, **11**, 796–807.

66. Shi,Y., Chan,D.W., Jung,S.Y., Malovannaya,A., Wang,Y. and Qin,J. (2011) A data set of human endogenous protein ubiquitination sites. *Mol. Cell. Proteomics*, **10**, M110 002089.

67. Ricciotti,E. and FitzGerald,G.A. (2011) Prostaglandins and inflammation. *Arterioscler. Thromb. Vasc. Biol.*, **31**, 986–1000.

68. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.

69. Croft,D., Mundo,A.F., Haw,R., Milacic,M., Weiser,J., Wu,G., Caudy,M., Garapati,P., Gillespie,M., Kamdar,M.R. *et al.* (2014) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **42**, D472–D477.

70. Kerrien,S., Orchard,S., Montecchi-Palazzi,L., Aranda,B., Quinn,A.F., Vinod,N., Bader,G.D., Xenarios,I., Wojcik,J., Sherman,D. *et al.* (2007) Broadening the horizon–level 2.5 of the HUPO-PSI format for molecular interactions. *BMC Biol.*, **5**, 44.

71. Roux,K.J., Kim,D.I., Raida,M. and Burke,B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, **196**, 801–810.

72. Inglis,D.O., Arnaud,M.B., Binkley,J., Shah,P., Skrzypek,M.S., Wymore,F., Binkley,G., Miyasato,S.R., Simison,M. and Sherlock,G. (2012) The Candida genome database incorporates multiple Candida species: multispecies search and analysis tools with curated gene and protein information for Candida albicans and Candida glabrata. *Nucleic Acids Res.*, **40**, D667–D674.

73. Lamesch,P., Berardini,T.Z., Li,D., Swarbreck,D., Wilks,C., Sasidharan,R., Muller,R., Dreher,K., Alexander,D.L., Garcia-Hernandez,M. *et al.* (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.*, **40**, D1202–D1210.

74. Drees,B.L., Thorsson,V., Carter,G.W., Rives,A.W., Raymond,M.Z., Avila-Campillo,I., Shannon,P. and Galitski,T. (2005) Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.*, **6**, R38.

75. Mani,R., St Onge,R.P., Hartman,J.L.T., Giaever,G. and Roth,F.P. (2008) Defining genetic interaction. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 3461–3466.

76. Baryshnikova,A., Costanzo,M., Myers,C.L., Andrews,B. and Boone,C. (2013) Genetic interaction networks: toward an understanding of heritability. *Annu. Rev. Genomics Hum. Genet.*, **14**, 111–133.

77. Schriml,L.M., Arze,C., Nadendla,S., Chang,Y.W., Mazaitis,M., Felix,V., Feng,G. and Kibbe,W.A. (2012) Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.*, **40**, D940–D946.

78. Osborne,J.D., Flatow,J., Holko,M., Lin,S.M., Kibbe,W.A., Zhu,L.J., Danila,M.I., Feng,G. and Chisholm,R.L. (2009) Annotating the human genome with Disease Ontology. *BMC Genomics*, **10(Suppl. 1)**, S6.

79. Mungall,C.J., Torniai,C., Gkoutos,G.V., Lewis,S.E. and Haendel,M.A. (2012) Uberon, an integrative multi-species anatomy ontology. *Genome Biol.*, **13**, R5.

80. Murali,T., Pacifico,S., Yu,J., Guest,S., Roberts,G.G. 3rd and Finley,R.L. Jr. (2011) DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Res.*, **39**, D736–D743.

81. Lardenois,A., Gattiker,A., Collin,O., Chalmel,F. and Primig,M. (2010) GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database (Oxford)*, **2010**, baq030.

82. Razick,S., Magklaras,G. and Donaldson,I.M. (2008) iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics*, **9**, 405.

83. Zuberi,K., Franz,M., Rodriguez,H., Montojo,J., Lopes,C.T., Bader,G.D. and Morris,Q. (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res.*, **41**, W115–W122.

84. Winter,A.G., Wildenhain,J. and Tyers,M. (2011) BioGRID REST Service, BiogridPlugin2 and BioGRID WebGraph: new tools for access to interaction data at BioGRID. *Bioinformatics*, **27**, 1043–1044.

85. Liu,G., Zhang,J., Choi,H., Lambert,J.P., Srikumar,T., Larsen,B., Nesvizhskii,A.I., Raught,B., Tyers,M. and Gingras,A.C. (2012) Using ProHits to store, annotate, and analyze affinity purification-mass spectrometry (AP-MS) data. Curr. Protoc. Bioinformatics , **Chapter 8**, Unit 8.16.

86. Liu,G., Zhang,J., Larsen,B., Stark,C., Breitkreutz,A., Lin,Z.Y., Breitkreutz,B.J., Ding,Y., Colwill,K., Pasculescu,A. *et al.* (2010) ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nat. Biotechnol.*, **28**, 1015–1017.

87. Aranda,B., Blankenburg,H., Kerrien,S., Brinkman,F.S., Ceol,A., Chautard,E., Dana,J.M., De Las Rivas,J., Dumousseau,M., Galeota,E. *et al.* (2011) PSICQUIC and PSISCORE: accessing and scoring molecular interactions. *Nat. Methods*, **8**, 528–529.